Business Intelligence - Introduction & Data Modeling (Pages 1-12)

This section provides a **detailed explanation**, followed by **quiz questions (MCQs & short-answer)** with **answers**.

📌 DETAILED EXPLANATION (PAGES 1-12)

1. Introduction to Business Intelligence (BI)

Definition:
Business Intelligence (BI) consists of **concepts, methods, and technologies** used to support business decision-making using **fact-based data analysis**.

Howard Dresner (Gartner Group) defines BI as:
"Concepts and methods to improve business decision-making by using fact-based support systems."

Key Characteristics of BI:
✓ **Decision Support** – BI helps businesses analyze data to improve operations.
✓ **Technology-Driven** – Uses specialized tools for data processing, storage, and visualization.
✓ **Data-Driven** – Converts raw data into actionable insights.
✓ **Continuous Evolution** – Includes traditional reporting, data warehouses, and AI-powered analytics.

---

2. Business Intelligence Journey

BI has evolved from **basic reporting systems** to **advanced AI-powered analytics**.

Milestones in BI Evolution:

1. **Basic Reporting (1980s):** Manual reports generated using spreadsheets.
2. **Classic BI (1990s-2000s):** Introduction of **data warehouses (DWs)** and SQL-based analytics.
3. **Big Data Era (2010s):** Rise of **cloud computing, NoSQL databases, real-time data processing**.
4. **AI & Advanced Analytics (2020s):** Machine learning (ML) and artificial intelligence (AI) integration.

---

3. Classic BI & Its Boundaries

BI is different from operational systems (**OLTP**) because it focuses on **analysis** rather than transaction processing.

Feature	Operational Systems (OLTP)	Business Intelligence (OLAP)
Purpose	Real-time transaction processing	Data analysis & decision-making
Data Storage	Live, transactional data	Historical, aggregated data
Query Speed	Optimized for small transactions	Optimized for large, complex queries
Data Model	Normalized (reduces redundancy)	Denormalized (improves query performance)

Challenges of Classic BI:

✓ **Complex Queries:** Requires multiple table joins for reports.

✓ **Slow Performance:** Large datasets slow down OLTP systems.

✓ **Scalability Issues:** Traditional BI systems struggle with big data.

---

4. Data Modeling for BI

BI systems use **structured data models** to improve performance and usability.

Normalization vs. Denormalization

Approach	Definition	Use Case
Normalization	Reduces data redundancy by dividing data into multiple tables	Used in **OLTP systems** for fast updates
Denormalization	Reduces table joins by combining data into fewer tables	Used in **BI/OLAP systems** for fast queries

Star Schema (BI Data Model)

The **Star Schema** is a simple BI data model used in **data warehouses**.

Components:

✓ **Fact Table** – Stores numerical business data (e.g., sales revenue, profit).

✓ **Dimension Tables** – Store descriptive attributes (e.g., product name, customer region).

Example Star Schema for Sales Data:

Fact Table: Sales	**Dimension Tables**
date_id, product_id, revenue	Date (date_id, year, quarter)
customer_id, store_id	Product (product_id, category)
region_id	Customer (customer_id, location)

Benefits of Star Schema:

✓ **Improves Query Performance** – Fewer joins speed up BI queries.

✓ **Easier for Business Users** – Simple structure allows non-technical users to explore data.

✓ **Scalability** – Handles large datasets efficiently.

---

5. Event Granularity in BI

Granularity refers to the **level of detail** in stored data.

Types of Granularity:

Granularity Level	**Example**	**Benefits**	**Drawbacks**
High Granularity	Each row represents a single transaction	More detailed analysis	Requires more storage
Medium Granularity	Each row represents an order summary	Balance between detail & storage	Some data is lost
Low Granularity	Each row represents total daily sales	Faster queries	No detailed insights

✓ Choosing the right granularity depends on business requirements and available data.

📌 QUIZ QUESTIONS & ANSWERS (PAGES 1-12)

✅ Multiple Choice Questions (MCQs)

1. **What is the primary goal of Business Intelligence?**
   a) To replace operational databases
   b) To support business decision-making using fact-based systems ✅
   c) To store large amounts of raw data
   d) To automate all business processes

2. **Which of the following best describes BI?**
   a) A technology that replaces human decision-making
   b) A system used only by IT teams
   c) A set of tools and methods for analyzing business data ✅
   d) A type of transaction processing system

3. **Why do organizations use a data warehouse for BI instead of operational databases?**
   a) Data warehouses process transactions faster
   b) Data warehouses are optimized for analytical queries ✅
   c) Operational databases store only historical data
   d) BI queries improve the performance of OLTP systems

4. **Which of the following is NOT a characteristic of a BI system?**
   a) Data-driven decision-making
   b) Uses structured and unstructured data
   c) Performs real-time transaction processing ✅
   d) Requires data integration from multiple sources

5. **What is the main advantage of a Star Schema in BI?**
   a) Improves query performance by reducing joins ✅
   b) Reduces data redundancy to zero
   c) Uses highly normalized tables
   d) Eliminates the need for data warehouses

✅ Short Answer Questions with Answers

1. **Define Business Intelligence (BI) in simple terms.**
   Answer: BI is the use of data, analytics, and technology to support business decision-making.

2. **What are two key benefits of BI for organizations?**
   Answer:
   ✓ Helps businesses make **better decisions**.
   ✓ Identifies **trends and patterns** for competitive advantage.
3. **What is the difference between OLTP and OLAP systems?**
   Answer:
   - **OLTP (Operational Systems)** → Handles real-time transactions (e.g., sales orders).
   - **OLAP (BI Systems)** → Used for data analysis and reporting (e.g., sales trends).
4. **Why is a Star Schema commonly used in BI?**
   Answer: It simplifies data querying by **reducing the number of joins**, improving performance.
5. **What is event granularity, and why is it important in BI?**
   Answer: Event granularity defines the **level of detail** at which business events are recorded. High granularity **captures more detail** but requires more storage, while low granularity **aggregates data for faster queries**.

✅ **Key Takeaways from Pages 1-12**

✓ BI supports business decision-making through data analysis.
✓ Data warehouses store structured, historical data optimized for BI queries.
✓ Star Schema improves query performance by reducing joins.
✓ Event granularity affects data storage and analysis depth.

Business Intelligence - Introduction & Data Modeling (Pages 13-24)

This section provides a **detailed explanation**, followed by **quiz questions (MCQs & short-answer)** with **answers**.

📌 DETAILED EXPLANATION (PAGES 13-24)

6. Data Warehouse & Analytical Data Modeling

A **data warehouse (DW)** is a centralized system designed to store and analyze large volumes of structured data.

Characteristics of a Data Warehouse:
✓ **Optimized for querying & analysis.**
✓ **Stores historical data for trend analysis.**
✓ **Integrates data from multiple sources.**

Why Can't We Use Operational Systems for BI?

Operational systems (OLTP) are **not suited for analytics** because:
✓ They focus on **real-time transactions**, not reporting.
✓ Queries on OLTP databases **slow down transactions**.
✓ OLTP systems **do not store historical data** for analysis.

Key Features of a Data Warehouse

✓ **Optimized for reading large data sets.**
✓ **Denormalized schema (Star Schema) improves query speed.**
✓ **Historical tracking** allows past trends to be analyzed.

---

7. Main Requirements for a BI Data Model

A **BI Data Model** should:
✓ Be **easy to understand** (so business users can query data easily).
✓ **Support fast queries** for large datasets.
✓ Store **historical data** for long-term analysis.

How Should We Structure Data for Analysis?

Requirement	Traditional OLTP (Operational)	BI Data Model (OLAP)
Data Consistency	Ensures real-time accuracy	Accepts slight delays for better performance
Query Performance	Optimized for small transactions	Optimized for large analytical queries
Data Storage	Uses normalized tables	Uses denormalized tables (Star Schema)

✓ **BI models use denormalization for better performance and ease of use.**

---

8. Star Schema in a Data Warehouse

A **Star Schema** is a simple data model used in BI systems.

Star Schema Components:

✓ **Fact Table** – Stores numerical business data (e.g., sales revenue, profit).
✓ **Dimension Tables** – Store descriptive data (e.g., product name, region, time).

✓ **Example Star Schema for Sales Data:**

Fact Table: Sales	Dimension Tables
date_id, product_id, revenue	Date (date_id, year, quarter)
customer_id, store_id	Product (product_id, category)
region_id	Customer (customer_id, location)

✓ **Benefits of Star Schema:**
✓ Improves **query performance** by reducing joins.
✓ Makes data **easy to navigate** for business users.
✓ Optimized for **large-scale data analysis**.

---

9. Event Granularity in BI

Granularity defines the **level of detail** at which data is stored in BI.

✓ **High Granularity** → Each row represents a single transaction.
✓ **Low Granularity** → Each row represents aggregated data (e.g., daily sales).

Choosing the Right Granularity

✓ **More detail allows deeper analysis** but increases storage needs.
✓ **Aggregated data speeds up queries** but may lose important insights.

✓ **Example:**

Granularity Level	Example
High Granularity	Each transaction (e.g., every product sold).
Medium Granularity	Each order (e.g., total revenue per order).
Low Granularity	Total daily sales (e.g., revenue per day).

✓ **The right granularity depends on the type of analysis needed.**

📌 **QUIZ QUESTIONS & ANSWERS (PAGES 13-24)**

✅ **Multiple Choice Questions (MCQs)**

1. **What is the primary purpose of a data warehouse?**
   a) To store real-time transactions
   b) To support business analysis and decision-making ✅
   c) To replace all operational databases
   d) To process customer orders
2. **Which of the following best describes a Star Schema?**
   a) A schema optimized for operational transactions
   b) A data model with one fact table and multiple dimension tables ✅
   c) A data model with only one table
   d) A schema that only stores normalized data
3. **Why are operational databases NOT suitable for BI analysis?**
   a) They store too much data
   b) They do not retain historical data ✅
   c) They are too simple
   d) They are optimized for business intelligence

4. **What does granularity in BI refer to?**
   a) The number of users accessing the database
   b) The level of detail in stored data ✅
   c) The process of cleaning data
   d) The total storage space used by a database

5. **Which factor is NOT considered when designing a BI data model?**
   a) Data performance
   b) User understandability
   c) Storage costs
   d) Customer satisfaction ✅

✅ Short Answer Questions with Answers

1. **What are the key benefits of using a data warehouse in BI?**
   Answer:
   ✓ Stores **historical data** for trend analysis.
   ✓ Improves **query performance** for analytical workloads.
   ✓ Integrates **data from multiple sources** for better decision-making.

2. **What is the difference between an OLTP system and a data warehouse?**
   Answer:
   - **OLTP System:** Stores **real-time transactional data**, optimized for **fast transactions**.
   - **Data Warehouse:** Stores **historical and aggregated data**, optimized for **queries & analysis**.

3. **What is a Star Schema, and why is it used in BI?**
   Answer: A Star Schema is a **BI data model** where a **central fact table** is linked to **multiple dimension tables**. It improves **query speed** and makes **data easier to explore**.

4. **Why is granularity important in BI?**
   Answer: Granularity determines the **level of detail** in stored data. High granularity provides **more detailed analysis**, while low granularity **improves query performance** by aggregating data.

5. **What are the advantages of denormalization in BI?**
   Answer:
   ✓ **Improves query performance** by reducing table joins.
   ✓ **Simplifies data retrieval** for business users.
   ✓ **Optimized for large-scale data analysis**.

✅ **Key Takeaways from Pages 13-24**

✓ A data warehouse stores historical, structured data optimized for BI analysis.
✓ A Star Schema improves query performance by reducing joins.
✓ Granularity affects data storage and analysis depth.
✓ Operational systems (OLTP) should not be used for BI queries.

Business Intelligence - Introduction & Data Modeling (Pages 25-36)

This section provides a **detailed explanation**, followed by **quiz questions (MCQs & short-answer)** with **answers**.

📌 DETAILED EXPLANATION (PAGES 25-36)

10. Analytical Data Models in BI

An **analytical data model** is a structured approach to storing and organizing data in a **data warehouse**.
It allows BI tools to **query and analyze data efficiently**.

Key Assumptions for Analytical Data Models:
✓ **Usability:** Business users should be able to access and analyze data easily.
✓ **Performance:** The model must support **fast queries on large datasets**.

Operational vs. Analytical Data Models

Feature	Operational Systems (OLTP)	BI Systems (OLAP/Data Warehouse)
Purpose	Real-time transaction processing	Business analysis & decision support
Data Structure	Highly normalized	Denormalized for query speed
Performance Focus	Fast inserts/updates	Optimized for large queries
Data Retention	Minimal historical data	Long-term historical storage

---

11. Data Warehouse as a Business Analysis Tool

A **data warehouse** acts as a **centralized data repository** to support BI tools.

Why Organizations Use Data Warehouses:
✓ **Faster Query Performance** – Queries are optimized for **big data analysis**.
✓ **Integrated View of Data** – Combines multiple operational systems.
✓ **Historical Tracking** – Stores past data for **trend analysis**.
✓ **Decision Support** – Provides structured, high-quality data for BI reports.

---

12. Data Modeling Techniques in BI

Data modeling determines **how data is structured in a warehouse**.

Star Schema (Most Common BI Model)

- **One fact table at the center** with multiple **dimension tables**.
- **Simplifies queries** and improves performance.
- **Example: Sales Data Model**
    - **Fact Table:** Sales (date_id, product_id, revenue, customer_id).
    - **Dimension Tables:**
      ✓ Product (product_id, category, brand).
      ✓ Customer (customer_id, location, age).
      ✓ Date (date_id, year, month, quarter).

---

13. Event Granularity in BI

Granularity defines **the level of detail at which data is stored in a fact table**.

Types of Granularity in BI Systems:

Granularity Level	Example	Pros	Cons
High Granularity	One row per **transaction**	Detailed analysis possible	Requires large storage
Medium Granularity	One row per **order**	Balances storage & detail	Some loss of data insight
Low Granularity	One row per **day/week**	Improves query speed	Limited data details

✓ Choosing the right granularity level depends on the type of analysis needed.

📌 **QUIZ QUESTIONS & ANSWERS (PAGES 25-36)**

✅ **Multiple Choice Questions (MCQs)**

1. **What is the main goal of an analytical data model?**
   a) To store raw unstructured data
   b) To improve data retrieval and analysis performance ✅
   c) To replace operational systems
   d) To perform real-time transactions

2. **Which is a key characteristic of a data warehouse?**
   a) Stores only real-time data
   b) Focuses on fast analytical queries ✅
   c) Replaces all operational databases
   d) Uses only normalized data

3. **Which of the following best describes a Star Schema?**
   a) A schema optimized for operational transactions
   b) A schema with one fact table and multiple dimension tables ✅
   c) A schema that does not support data analysis
   d) A schema with only one table

4. **Why is granularity important in BI?**
   a) It defines the data structure of a star schema
   b) It determines the level of detail in a fact table ✅
   c) It limits the number of queries
   d) It eliminates data redundancy

5. **What is one key reason businesses use a data warehouse?**
   a) To optimize real-time transactions
   b) To analyze and store historical data ✅
   c) To reduce data volume
   d) To eliminate business intelligence

✅ Short Answer Questions with Answers

1. **What is the difference between an operational system and a data warehouse?**
   Answer:
   - **Operational System (OLTP):** Handles real-time transactions, optimized for quick updates.
   - **Data Warehouse (OLAP):** Stores historical and aggregated data, optimized for analysis and reporting.

2. **What is a Star Schema, and why is it used in BI?**
   Answer: A Star Schema is a **data model** where a **central fact table** is linked to **multiple dimension tables**. It is used in BI because it **improves query speed** and makes data **easier to explore**.
3. **Why do businesses use data warehouses for BI instead of operational systems?**
   Answer:
   ✓ Data warehouses store **historical data** for analysis.
   ✓ They improve **query performance** for large datasets.
   ✓ They integrate data from **multiple sources**.
4. **What is event granularity in BI?**
   Answer: Event granularity refers to the **level of detail** at which business events are recorded in a fact table. High granularity stores **individual transactions**, while low granularity **aggregates data** (e.g., daily sales totals).
5. **What are the advantages of using a data warehouse for business analysis?**
   Answer:
   ✓ **Faster data access** – Optimized for analytical queries.
   ✓ **Integrated data** – Combines multiple data sources.
   ✓ **Historical trend tracking** – Allows businesses to analyze past performance.

✅ **Key Takeaways from Pages 25-36**

✓ Analytical data models improve BI performance and usability.
✓ Data warehouses store structured, historical data optimized for BI analysis.
✓ Star Schema simplifies querying by reducing joins.
✓ Granularity affects data storage, query performance, and analysis depth.

Business Intelligence - Introduction & Data Modeling (Pages 37-48)

This section provides a **detailed explanation**, followed by **quiz questions (MCQs & short-answer)** with **answers**.

📌 DETAILED EXPLANATION (PAGES 37-48)

14. Why Businesses Need a Data Warehouse

A **data warehouse (DW)** is essential for business intelligence because **operational systems (OLTP) are not designed for analytics**.

Key Reasons for Using a Data Warehouse:

✓ **Faster Data Access** – BI queries run faster than in operational systems.
✓ **Integrated Data** – Combines multiple sources for a **holistic business view**.
✓ **Historical Tracking** – Stores data over time for **trend analysis**.
✓ **Improved Decision-Making** – Provides **accurate, structured** insights.

Challenges of Using Operational Databases for BI:

- **OLTP systems handle transactions**, not large-scale analysis.
- **Queries on OLTP databases slow down business operations.**
- **Data from different systems may not be standardized.**

---

15. Ensuring Data Consistency in a Data Warehouse

A **BI system must ensure that all reports and dashboards display consistent data**.

How Data Warehouses Ensure Data Consistency:

✓ **ETL Process** (Extract, Transform, Load) ensures **data cleaning & standardization**.
✓ **Business Rules & Data Validation** – Ensures **data accuracy & integrity**.
✓ **Single Source of Truth (SSOT)** – All departments use the **same KPI definitions**.

Example of Data Inconsistency:

- Finance and Sales define **revenue differently**, causing reporting discrepancies.
  ✓ **Solution:** A Data Warehouse ensures **consistent KPI calculations**.

---

16. Event Granularity in BI

Granularity defines the **level of detail in a fact table**.

Granularity Level	Example	Benefits	Drawbacks
High Granularity	Each row represents **one transaction**	Detailed analysis possible	High storage requirement
Medium Granularity	Each row represents **an order summary**	Balanced detail & storage	Some loss of insights
Low Granularity	Each row represents **daily totals**	Faster query performance	Limited insights

✓ Choosing the right granularity level depends on business needs.

---

17. Types of Fact Tables in BI

Fact tables store **measurable business data**. There are **three main types**:

1. Transactional Fact Tables

- Stores **individual business events**.
- **Example:** Sales transactions (date, product, quantity, revenue).
- **Use case:** Best for tracking **detailed** business events.

2. Periodic Snapshot Fact Tables

- Stores **data snapshots** at regular intervals (**daily, weekly, monthly**).
- **Example:** Monthly account balances or warehouse inventory levels.
- **Use case:** Best for tracking **performance trends** over time.

3. Accumulating Snapshot Fact Tables

- Tracks the **lifecycle of a business process**.
- **Example:** Order processing (order placed → shipped → delivered → paid).
- **Use case:** Best for tracking **multi-stage business processes**.

✓ Transactional facts store raw events, periodic snapshots track trends, and accumulating snapshots follow process lifecycles.

📌 QUIZ QUESTIONS & ANSWERS (PAGES 37-48)

✅ Multiple Choice Questions (MCQs)

1. **What is the primary reason businesses use a data warehouse?**
   a) To optimize real-time transactions
   b) To analyze and store historical data ✅
   c) To reduce data volume
   d) To eliminate business intelligence

2. **Which of the following ensures data consistency in a data warehouse?**
   a) Storing all data in operational databases
   b) Using different KPI definitions across departments
   c) Implementing ETL processes ✅
   d) Ignoring historical data

3. **What does event granularity in BI refer to?**
   a) The total number of records in a database
   b) The level of detail in stored data ✅
   c) The number of database joins
   d) The speed of data processing

4. **Which type of fact table is best for tracking multi-stage business processes?**
   a) Transactional Fact Table
   b) Periodic Snapshot Fact Table
   c) Accumulating Snapshot Fact Table ✅
   d) OLTP Fact Table

5. **Which is NOT a characteristic of a data warehouse?**
   a) Stores historical data
   b) Optimized for analytical queries
   c) Designed for real-time transaction processing ✅
   d) Integrates data from multiple sources

✅ Short Answer Questions with Answers

1. **Why do businesses use a data warehouse for BI instead of operational systems?**
   Answer:
   ✓ Data warehouses store **historical data** for analysis.
   ✓ They improve **query performance** for large datasets.
   ✓ They integrate data from **multiple sources** for a **holistic business view**.
2. **What is event granularity in BI?**
   Answer: Granularity refers to the **level of detail** in stored business data. High granularity provides **detailed transaction-level data**, while low granularity **aggregates data** for performance efficiency.
3. **What are the three types of fact tables in BI?**
   Answer:
   - **Transactional Fact Tables** – Store **individual transactions** (e.g., each product sale).
   - **Periodic Snapshot Fact Tables** – Store **aggregated data over time** (e.g., monthly balances).
   - **Accumulating Snapshot Fact Tables** – Track **multi-step business processes** (e.g., order processing).
4. **How does ETL ensure data consistency in a BI system?**
   Answer: The ETL process **extracts data from operational systems**, cleans and standardizes it, and loads it into a **data warehouse**, ensuring **data integrity and consistency**.
5. **Why is maintaining a Single Source of Truth (SSOT) important in BI?**
   Answer: SSOT ensures that **all departments use the same data definitions** and KPIs, avoiding inconsistencies in reports.

✅ **Key Takeaways from Pages 37-48**

✓ A data warehouse ensures fast, integrated, and accurate BI reporting.
✓ ETL processes clean, standardize, and load data into a warehouse.
✓ Fact tables store numerical business data and can be transactional, periodic, or accumulating.
✓ Event granularity affects storage and query performance.

Business Intelligence - Introduction & Data Modeling (Pages 49-60)

This section provides a **detailed explanation**, followed by **quiz questions (MCQs & short-answer)** with **answers**.

📌 DETAILED EXPLANATION (PAGES 49-60)

18. Types of Facts Based on Additivity

Fact tables store **measurable business data**, but not all facts behave the same when aggregated.

Three Types of Facts Based on Additivity:

Fact Type	Definition	Example
Fully Additive	Can be summed across **all dimensions**.	Total revenue, total quantity sold.
Semi-Additive	Can be summed **for some dimensions, but not all**.	Account balance (can be summed over regions but not over time).
Non-Additive	Cannot be summed at all; requires other aggregation methods.	Unit price, percentages (e.g., profit margin).

✓ Fully additive facts are the easiest to use, while non-additive facts require careful handling in reports.

---

19. The ETL Process (Extract, Transform, Load)

BI systems rely on **ETL (Extract, Transform, Load)** to move and prepare data for analysis.

Steps in ETL:

✓ **Extract:** Pulls data from operational systems (e.g., sales databases, CRM).
✓ **Transform:** Cleans, standardizes, and formats data (e.g., converting currencies, removing duplicates).
✓ **Load:** Moves transformed data into the **data warehouse** for analysis.

✓ ETL ensures that BI data is accurate, consistent, and up-to-date.

Challenges of ETL:

- **Data Quality Issues** – Inconsistent data formats need cleaning.
- **Processing Time** – Large datasets take time to transform.
- **Data Latency** – ETL is not real-time; updates occur in batches.

---

20. Slowly Changing Dimensions (SCDs)

What Are Slowly Changing Dimensions?

- **Dimensions (e.g., customers, products) do not change often but need version control.**
- **Example:** A customer moves to a new city. Do we keep the old location or overwrite it?

Types of SCDs:

SCD Type	Description	Example
Type 0	No changes allowed (fixed values).	Customer registration date.
Type 1	Overwrites old values.	Customer phone number update.
Type 2	Creates a new row for the updated record.	Customer changes address.

✓ Type 2 SCDs track historical changes, while Type 1 simply updates records.

---

21. Surrogate Keys in BI

What Is a Surrogate Key?

- **A surrogate key (SK) is an artificial, system-generated identifier** (e.g., auto-incremented ID).
- Used in **dimension tables instead of natural keys** (e.g., customer IDs from CRM).

✓ Surrogate keys improve performance, ensure uniqueness, and support historical tracking (SCDs).

---

22. Schema Types in Data Warehouses

BI data warehouses use different **schemas** (data structures) to organize fact and dimension tables.

1. Star Schema

- **Most common BI schema** for fast querying.
- **One central fact table** connects to multiple **dimension tables**.

✓ Fast queries but some data redundancy.

2. Snowflake Schema

- **More normalized than a star schema** (dimension tables are broken down into smaller tables).
- **Example:** A **"Product" table** is split into "Product Category" and "Product Details".

✓ Reduces redundancy but requires more joins, slowing down queries.

📌 **QUIZ QUESTIONS & ANSWERS (PAGES 49-60)**

✅ **Multiple Choice Questions (MCQs)**

1. **What is an example of a fully additive fact?**
   a) Profit margin
   b) Total revenue ✅
   c) Unit price
   d) Customer satisfaction score
2. **Why do BI systems use ETL processes?**
   a) To generate AI models
   b) To move and transform data before loading it into a data warehouse ✅
   c) To delete old records
   d) To perform live data transactions
3. **What is the difference between Type 1 and Type 2 Slowly Changing Dimensions (SCDs)?**
   a) Type 1 tracks historical changes, Type 2 overwrites data
   b) Type 1 overwrites data, Type 2 tracks historical changes ✅
   c) Both overwrite historical data
   d) Both create new records for every update

4. **Why are surrogate keys used in BI?**
   a) To store natural business keys
   b) To ensure uniqueness and improve performance ✅
   c) To replace data transformations
   d) To store customer names

5. **Which schema type is **most commonly used in BI**?
   a) Star Schema ✅
   b) Snowflake Schema
   c) OLTP Schema
   d) Hybrid Schema

✅ Short Answer Questions with Answers

1. **What is a fully additive fact in BI?**
   Answer: A fact that can be summed across **all dimensions**, such as **total revenue**.

2. **What are the three steps of the ETL process?**
   Answer:
   - **Extract** (collects raw data),
   - **Transform** (cleans and standardizes),
   - **Load** (moves data into the warehouse).

3. **Why do businesses use Slowly Changing Dimensions (SCDs)?**
   Answer: SCDs **track changes in dimension attributes over time** to maintain accurate historical records.

4. **What is the difference between Star Schema and Snowflake Schema?**
   Answer:
   - **Star Schema** has a **simple structure** with denormalized dimensions (fast queries).
   - **Snowflake Schema** is **more normalized** (reduces redundancy but requires more joins).

5. **Why are surrogate keys preferred over natural keys in BI?**
   Answer:
   - They **ensure uniqueness**.

- They **improve query performance**.
- They **support historical tracking (SCDs)**.

✅ **Key Takeaways from Pages 49-60**

✓ Fact tables store measurable business data and can be fully, semi, or non-additive.
✓ ETL processes move, clean, and transform data for BI systems.
✓ Slowly Changing Dimensions (SCDs) help track historical changes in dimension tables.
✓ Surrogate keys improve performance and data integrity.
✓ Star Schemas improve query speed, while Snowflake Schemas reduce data redundancy.